

Multiscale integration of contextual information during a naturalistic task

Article (Accepted Version)

Keidel, James, Oedekoven, Christiane, Tut, Andreea and Bird, Chris (2018) Multiscale integration of contextual information during a naturalistic task. *Cerebral Cortex*, 28 (10). pp. 3531-3539. ISSN 1047-3211

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/70366/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Multiscale integration of contextual information during a naturalistic task

James L. Keidel, Christiane S.H. Oedekoven, Andreea C. Tut and Chris M. Bird

University of Sussex

RUNNING TITLE: Multiscale integration of context

School of Psychology
University of Sussex
Falmer
BN1 9QH
United Kingdom

Corresponding author:
Dr James Keidel
School of Psychology
University of Sussex
Falmer
BN1 9QH
United Kingdom

E: j.keidel@sussex.ac.uk
T: +44 (0)1273 872 766
F: +44 (0)1273 678 058

Abstract

Everyday experience requires rapid and automatic integration of incoming stimuli with previously stored knowledge. Prior knowledge can help construct a general “situation model” of the event, as well as aid comprehension of an ongoing narrative. Using fMRI in healthy adult humans we investigated processing of videos whose locations and characters were always familiar but whose narratives were either a continuation or non-continuation of an earlier video (high context (HC) or low context (LC) respectively). Responses in parahippocampal gyrus and retrosplenial cortex were composed of an initial transient, locked to the video onsets, followed by a period of lower amplitude activation that was greater in the LC condition. This may reflect rapid processing of core components of situation models such as location and characters and more gradual incorporation of their narrative themes. By contrast, activity increases in left hemisphere middle temporal gyrus (MTG), angular gyrus and inferior frontal gyrus were maintained throughout the videos and were higher for HC versus LC videos. Further, activity in the left MTG peaked earlier in the HC condition. We suggest that these regions support representations of the specific inter-linked concepts necessary to comprehend an ongoing narrative, which are already established for the HC videos.

KEYWORDS: Conceptual knowledge; situation models; parahippocampal gyrus; retrosplenial cortex; semantic network

Introduction

Making sense of everyday life requires integration of information from multiple sources and at multiple scales. First, we need to perceive and construct a representation of the large-scale environment – where are we and who is here? Then we have to identify the broad themes of the situation—colloquially, “what’s going on”. Are the people talking to one another? What are they talking about? Do they like each other? Once we have established these basic facts we can begin to attend to the smaller-scale specifics of the conversation or the action, in order to understand what is happening moment by moment and update our representations accordingly.

As common as this kind of processing is, it is quite difficult to study it within the MRI scanner for obvious logistical reasons. Though the use of naturalistic video stimuli has proven effective in circumventing some of the associated problems (Zacks et al. 2001; Hasson et al. 2008; van Kesteren et al. 2010; Bird et al. 2015; St-Laurent et al. 2015), one limitation of the use of such stimuli in fMRI experiments is the difficulty in creating matched sets over which to perform contrasts. The inherent complexity of naturalistic videos and the many dimensions across which they may differ render it difficult to confidently ascribe differences in BOLD response to particular aspects of the stimulus. Further, a typical analysis strategy that we and others have used (Bird et al. 2015; Chen et al. 2016; Schlottermeier et al. 2016) involves modeling the entire ~50 s response to the video as an average response or with a single regressor. Such a model implicitly assumes that the response in a given voxel is more or less constant across the duration of the video. However, it is likely that different

regions have significantly different response profiles, and therefore a more flexible modeling strategy is needed to capture this finer-grained information.

In the current study we focus on how different brain regions work over different temporal and spatial scales to integrate previously acquired knowledge into the unfolding representation of a new, related video. On the one hand, there is the larger-scale and more slowly changing information that provides the basis for creation of a “situation model” describing the location, the characters and the main theme of the video. On the other hand is the processing of the rapidly changing details of the narrative as it unfolds.

To study these effects, we used a specially selected set of video stimuli taken from situational comedy series (henceforth “sitcoms”). Specifically, we obtained pairs of scenes from 20 different sitcoms in which the characters and backdrop were the same, but the topic of discussion was different. Participants saw the first half of one clip from each show (no context condition: NC) then saw either the matching second half of that clip (high context: HC) or the second half of the other clip from the same show (low context: LC). The crucial feature of this design is the ability to observe differences in neural responses to completely identical video stimuli driven solely by the degree of context provided by the NC video from the same sitcom. To investigate whether the time course of responses to the HC and LC videos differ, we compare the activity across each scan acquisition point during the first 26 seconds of the videos.

A situation model for a video includes features such as the location, the main protagonists and their intentions and motivations (Zwaan and Radvansky 1998). Ranganath and Ritchey (2012) have proposed that the parahippocampal gyrus (PHG) and retrosplenial cortex (RSC) are essential for constructing and

updating situation models, based on their established role in representing contextual associations (Bar and Aminoff 2003; Aminoff et al. 2013) and processing of large-scale space (Maguire 2001; Epstein 2008). We might therefore expect that responses within the PHG and RSC would be strongest immediately after the new scene begins, since information about large-scale characteristics of the scene is available immediately after the onset of the video. However, identifying the thematic content (and correspondingly, the intentions and motivations of the protagonists) may take time, particularly if this information emerges during an unfolding narrative. Consequently, processing in the PHG and RSC, corresponding to the establishment and updating of the situation model, is predicted to continue well into the course of the video.

In contrast to the establishment of situation models whose core components change only slowly during the videos, developing a more detailed understanding of the narrative requires continuous access to specific concepts and associations between them. This type of semantic processing is associated with left hemisphere regions such as the angular gyrus (AG), middle and inferior temporal gyri (MTG and ITG) and the inferior frontal gyrus (IFG) (Humphries et al. 2007; Binder et al. 2009). In our study, videos in the HC condition were straightforward continuations of a recently presented video, and therefore representations of specific details of the discussion had already been established. In the LC condition, however, no such prior representations were available. Although we expected all videos to elicit activity in regions associated with conceptual processing, we anticipated that the magnitude of this activity would be modulated by participants' previous exposure to the narrative.

Materials and Methods

Participants

Twenty-one healthy right-handed native English-speaking participants took part in this study (10 female, mean age 22.6). All participants gave informed consent under a protocol approved by the Research Governance and Ethics Committee of the University of Sussex, and were paid £15 for their participation.

Stimuli

A total of 80 video clips were used in this study. These clips were taken from 20 US and UK television series that aired between 1970-2003, selected to be unfamiliar to our participant group, with two clips from each show that were then split into first and second halves (see Supplementary Appendix for a list of shows and episodes). Clips were selected on the basis of having the exact same set of characters in the same room, so that the perceptual aspects of the stimuli were matched as closely as possible. The mean duration of the clips was 34.4 s (SD = 8 s). The first halves of clips (M = 29.2 s, SD = 5.2) were on average shorter than second halves (M = 39.7, SD = 7.53). Mean durations of first and second halves did not differ significantly across the lists.

Procedure

Before scanning, participants were instructed that they would see and hear video clips in sets of five. In the first set and subsequent odd sets, they would see the first half of an excerpt from an old sitcom. In the second set and subsequent even sets they would see the second half of a clip. Participants were asked to stay as still as possible and watch the clips just as they would watch television at

home. They were also informed that there would be a short memory test for details concerning the clips.

The experiment was composed of two runs of approximately 17 minutes each, with two alternating sets of five first halves and five second halves in each run (a total of 20 clips per run). Four running order lists were created such that $\frac{1}{4}$ of the participants saw each of the four possible orders of the two first halves and two second halves from each show (and therefore every clip was shown as both part of a matching and non-matching pair, with each participant seeing 40 clips in total). Each video was shown on a black background, with a white fixation cross presented for 16 s between each video. On average, there was a gap of about two minutes between the presentation of a first half and the second half from that same show.

After leaving the scanner, participants were brought to a testing room. They first filled out a questionnaire asking whether they were familiar with the shows they had seen. Only four participants reported familiarity with any of the shows, and these participants only recognized 1-3 of the shows presented, and in no case did they recognize the specific scene being watched. As this represents less than 2% of the stimuli used we were satisfied that as a group the participants were unfamiliar with the stimulus set. Participants then performed a memory task on a computer. Participants were shown the first 4-6 seconds of the second half of each video they saw and were then shown an open-ended question concerning the clip. For instance, the question for a scene in which a wife and husband discuss a cricket match was "At what University did Paul play cricket?" These questions varied in difficulty and specificity and were designed both to ensure participants paid attention to the stimuli as well as to provide

variability in the level of performance across participants. Pilot behavioral data indicated that participants could answer them reliably. They responded to the question by typing in an answer in a response field below the question.

MRI Acquisition

BOLD-sensitive T2*-weighted images were acquired during task performance on a 1.5 T Siemens Avanto MRI scanner using a gradient-echo EPI pulse sequence with the following parameters: FOV 192 x 192 mm, TR = 2.62, echo time 42 ms, flip angle 90 degrees, 35 ascending 3 mm slices, 0.6 mm gap. In addition, a high-resolution T1-weighted MP-RAGE structural image was acquired with the following parameters: FOV 256 x 256 mm, 1mm isotropic voxels, TR 2.73 s, TE 3.57 ms.

Image Preprocessing

All MRI processing was carried out in AFNI (Cox 1996). The first five volumes acquired in each functional run were discarded to allow for T1 equilibration. Images were then time-shifted to correct for offsets in acquisition time across slices within a volume. The affine transform and warps required to align all EPI volumes to the first EPI volume, to align the EPI to the anatomy and to normalize the data to Talairach space (Talairach and Tournoux 1988) were concatenated and applied to the EPI data simultaneously to reduce the number of interpolation steps required. The resulting EPI images were blurred using AFNI's 3dBlurInMask to have an approximate final FWHM of 9 mm, and scaled to a mean of 100 to convert to percent signal change.

GLM analysis

Block design first-level GLM analysis was carried out in AFNI's 3dDeconvolve, and included task regressors for NC videos (20 clips), LC videos (10 clips) and HC videos (10 clips). This analysis was used only to generate maps of activation for each condition vs. rest, mainly for comparison with previously reported results. We also carried out an analysis identical to the above, except that instead of modeling videos as blocks, we modeled only the onset of the video, using a gamma variate. Further, we carried out a block-level analysis as above, but collapsing across the two types of second halves (HC and LC) to allow us to quantify the degree of repetition suppression induced by repetition of a show generally. Finally, to analyze changes in activation over time throughout the videos, we used a set of task regressors modeling the three conditions instantiated as a set of 12 piecewise linear tent functions modeling the first 26.2 seconds of response from the onset of each video from a given condition. We chose this duration to model because the focus of our study was the difference between HC and LC videos, and the shortest second half clip was 25 s. Thus, by restricting our analysis to this timeframe we could ensure that we were only modeling responses to video, without mixing in any response to rest. In addition, the six motion parameters obtained during preprocessing and baseline Legendre polynomials up to the 9th degree were included as regressors of no interest. A series of group-level analyses were carried out via the AFNI program 3dANOVA3 –type 4, with time and condition as fixed factors and subject as a random factor. The first set of analyses investigated the entire 26.2 s period that we modeled, and thus included all time points except the first, as this by definition does not reflect processing of a video. For whole-brain significance correction, residual time series from each subject's first-level GLM were

analyzed with AFNI's 3dFWHMx using the mixed spatial autocorrelation option and averaged together, resulting in a smoothness estimate equivalent to 10 mm FWHM. This newer option in AFNI provides improved defense against Type 1 error as the estimated smoothness is greater than that based on a Gaussian autocorrelation function. The averaged ACF values ($a = 0.46$, $b = 4.06$, $c = 7.81$) were passed to AFNI's 3dClustSim, resulting in a necessary cluster size of 18 voxels at an uncorrected p of .001 using bi-sided thresholding, cluster formation requiring shared edges (nearest neighbor setting = 2), and a mask generated by resampling the TT_N27 brain supplied with AFNI onto a 3 mm³ grid to achieve a whole-brain corrected p of .05.

To analyze activation on a timepoint-by-timepoint basis within clusters identified in the whole-brain ANOVA while avoiding circular voxel selection, we employed the leave-one-subject out (LOSO) method described by Esterman (2010). In this method, the voxels selected for a given ROI in a given subject are determined by the whole-brain suprathreshold voxels identified in a whole-brain analysis carried out with all the other subjects in the study. Thus, for instance, the voxels chosen for the left MTG ROI in subject one were those identified in an ANOVA that excluded that subject. The same process was used for all other subjects and ROIs reported below.

RSA analysis

Preprocessing for the RSA analysis was similar to that for the GLM analysis, except that the raw data were not smoothed and were not normalized to a standard template. We carried out two RSA analyses. The first was designed to identify areas where clips from the same show, irrespective of condition, elicited similar multivariate patterns of activation. The second analysis directly

contrasted similarity between first halves and HC videos and first halves and NC videos. Surprisingly, and in contrast to previous findings using similar designs (Bird et al. 2015; St-Laurent et al. 2015; Chen et al. 2016), we observed few significant effects in these analyses. Further details of the methods and results for these analyses can be found in the Supplementary Material.

Results

Behavioral results

Participants answered an average of 62.4% of questions correctly. It is important to note all questions were open-ended and therefore this is a very high level of performance. Questions about HC videos were answered correctly 67.8% of the time and questions concerning LC videos were answered correctly 58.1% of the time. The difference of 9.7% was significant in a one-tailed t-test ($t(20) = 2.0, p = .029$).

fMRI results

One-sample t-tests indicated that videos in all three conditions elicited patterns of activation consistent with previous studies, including extensive activity in the thalamus as well as primary and secondary visual and auditory cortices, extending into the inferior and anterior temporal cortices respectively (Supplementary Figures 1-3).

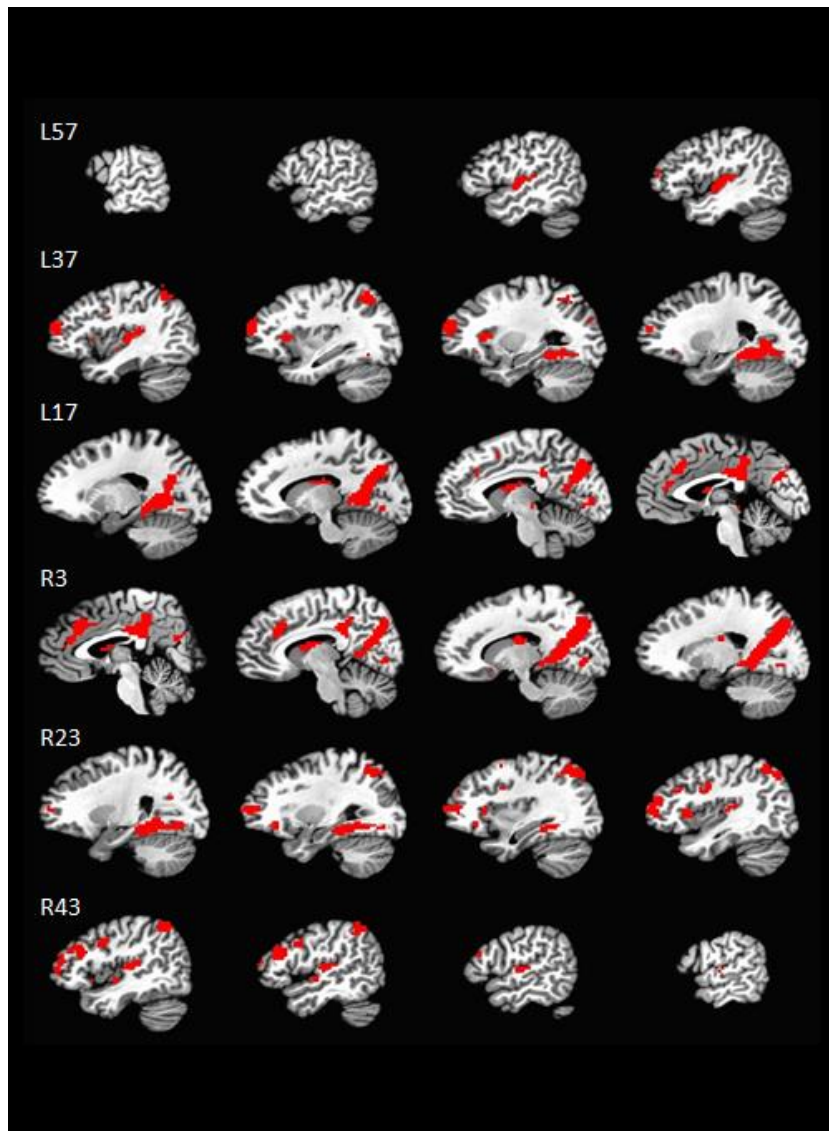


Figure 1. Conjunction map of positive onset responses to video clips. Voxels in red represent whole-brain significant responses ($p < .001$ uncorrected threshold) to the onset of a video clip in all three (NC, LC and HC) conditions.

Measurement of responses to the onset of the videos revealed a clear pattern of activation in the bilateral RSC and PHG, noticeably not including earlier areas of the visual stream. Figure 1 depicts a conjunction map of whole-brain corrected clusters showing greater activation in response to the onset of a video with respect to rest.

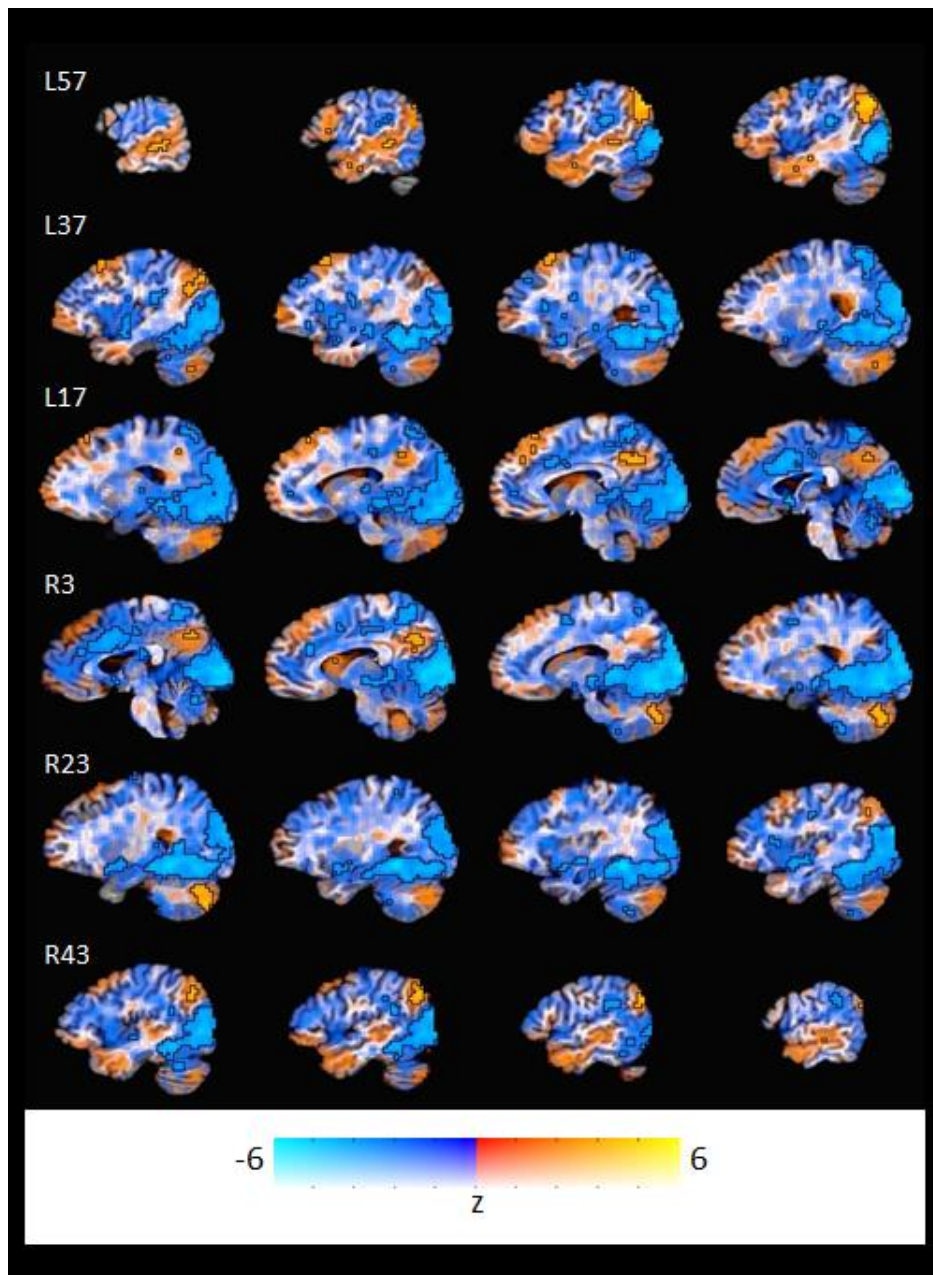


Figure 2. Contrast of first vs. second halves of videos using block regressors. Voxels in cool colors responded significantly less strongly to second halves, while voxels in warm colors responded more strongly. Beta coefficients are mapped to hue and t-statistics to transparency, following Allen et al. (2012). Voxels enclosed within black lines are significant at $p < .001$ uncorrected.

Contrasting responses to first and second halves of videos (i.e., collapsing over HC and LC) revealed extensive repetition suppression bilaterally throughout the ventral and dorsal visual streams (including PHG but not RSC), as well as in the middle and anterior cingulate gyri, the bilateral insula, the bilateral inferior parietal lobule and the right superior parietal lobule (Figure 2)

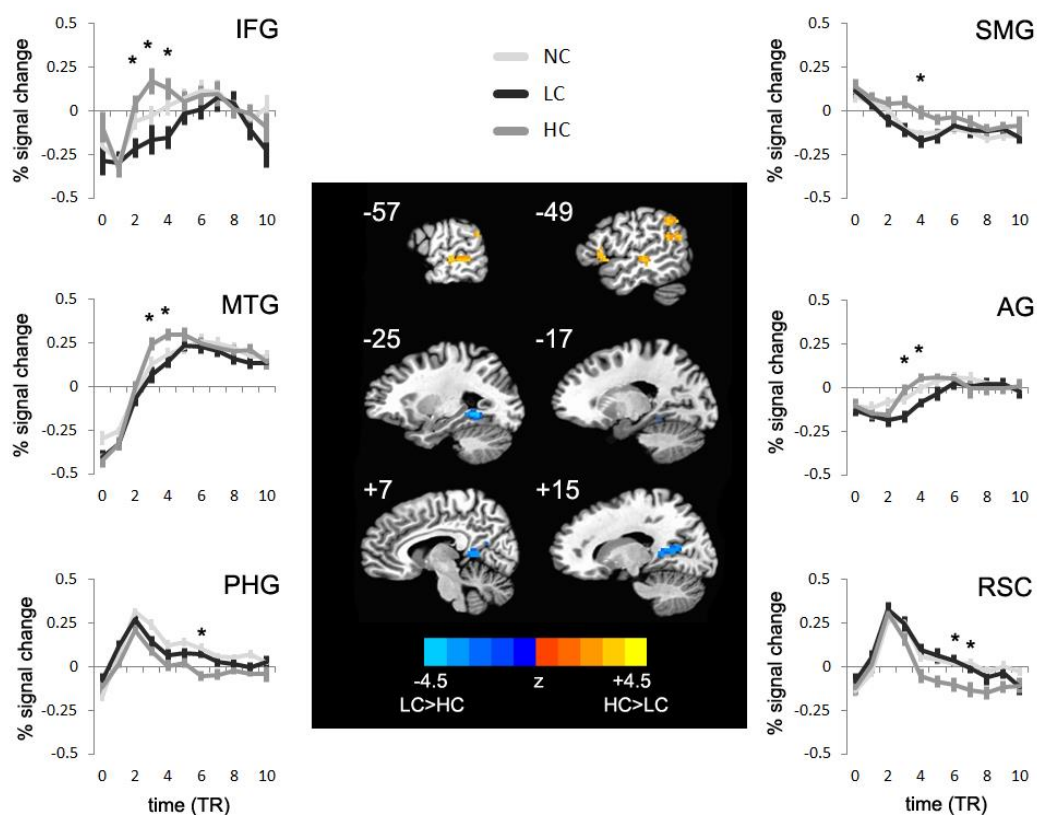


Figure 3. Main effect of condition in Condition x Time ANOVA. Clusters depicted in warm colors represent areas with greater mean activation for the HC than the LC condition, with cool colors denoting the opposite effect. Graphs represent percent signal change over time in the depicted clusters, with error bars representing +/1 standard error of the mean. Values in the graphs were calculated as the mean across voxels in each cluster, all of whose time courses

were scaled to be percentages of the mean signal across the run in each voxel. All time course estimates were generated using the leave-one-subject-out procedure described in the Methods.

To investigate the effect of prior contextual knowledge over the initial 26.2 s period, we contrasted BOLD activation in a 2 (Condition: LC and HC) x 10 (TR post-stimulus: 2-11) ANOVA. A significant main effect of Condition was observed in right RSC and left PHG, MTG, IFG, AG and supramarginal gyrus (SMG). Taking the mean activation in each condition across time, it is clear that the effects in MTG, IFG, AG, and SMG reflect a relative increase in BOLD activity for HC videos, while the effects in PHG and RSC reflect decreased activation for HC videos relative to LC (Figure 3). Further, a significant interaction between Condition and TR was observed in the left MTG. As can be seen in Figure 3, this interaction reflects a quicker rise and peak in the HC than in the LC condition. Notably, this MTG cluster overlaps strongly with the MTG cluster showing a main effect of Condition (Dice coefficient = .30—maximum Dice coefficient given unequal ROI sizes = .71—see Supplementary Figure 4).

To examine the time courses in more detail, we next tested the difference between HC and LC at each time point in each cluster. To avoid circularity, we used leave-one-subject-out (LOSO) cross-validation (Esterman et al. 2010). To correct for multiple comparisons, we employed a threshold of $p < .005$, based on the 10 tests in each region (i.e., $0.05/10$; a Bonferroni correction across all tests together would almost certainly be overly conservative given the dependency between tests). In the PHG, a significant difference between conditions was observed only 15.72 s post-stimulus, though response magnitude to HC videos

was lower than LC at all time points. In the RSC, the difference between HC and LC was significant from 10.48 s to 15.72 s. In the AG and MTG clusters identified in the analysis of the main effect of condition, significantly greater responses to HC videos were observed at 7.86 and 10.48 s post-stimulus. In the IFG, HC-elicited activation significantly surpassed LC activations at 5.24, 7.68 and 10.48 s post-stimulus. Finally, in the SMG a significant HC > LC effect was observed 10.48 s after stimulus onset. Overlap maps of the LOSO-identified clusters are provided in Supplementary Figure 5.

Discussion

In this study we compared the brain's activity during viewing of videos taken from sitcoms that were either straightforward continuations of an earlier clip (HC condition) or clips involving locations and characters that were familiar but where the narrative topic was new (LC condition). This enabled us to identify brain regions involved in the integration of prior knowledge with incoming sensory information during the processing of naturalistic stimuli. Responses in RSC and PHG consisted of a transient, time-locked peak to the videos' onsets, followed by a period of lower activation that was greater in the LC condition. We interpret this as reflecting the ongoing incorporation of novel thematic information into the situation model of the LC videos. By contrast, responses in left lateral temporal and parietal areas were generally slower to peak, were maintained throughout the videos, and were higher during the HC condition. We interpret this as resulting from activation and maintenance of linked conceptual information relevant for understanding the specific content of the narrative taking place in the videos.

Previous work suggests that processing in the PHG/RSC plays two different, though likely complementary roles: on the one hand, they play a central role in visuo-spatial processing of large-scale scenes (Epstein 2008; Vann et al. 2009; Mullally and Maguire 2011), while on the other they are sensitive to broader contextual or situational associations (Ranganath and Ritchey 2012; Aminoff et al. 2013). The robust activation that we observed in these regions in response to the onset of the videos (Figures 1 and 3) likely reflects processing of visuo-spatial information in the scene as well as recognition of the general context (e.g., a kitchen, a public house etc.). Although these responses are driven by visual input, they are very different to those seen in early visual regions where activity is sustained throughout the video (Supplementary Figure 6). The shape and timecourse of these onset responses (Figure 1), particularly in the HC condition, is highly similar to the hemodynamic response typically elicited in PHG by still images of large-scale spatial structures (Epstein and Kanwisher 1998; Epstein et al. 2003). Similar transient responses to the presentation of indoor and outdoor scenes within a continuous video were reported by Hasson et al. (2004).

The critical differences between the HC and LC conditions occur after this early peak in activation, around 16s into the videos (Figure 3). This is unlikely to be driven by processing of the scene information and may instead reflect differences between the broader situational content of the conditions. A situation model includes information such as the location, the protagonists present and the thematic content of the event. While the location and characters are identical in our HC and LC conditions, the thematic content is changed and we argue that this drives the increased activation observed in PHG/RSC in the LC

condition. An appreciation of the main themes of the videos requires an understanding of what the characters are talking about. This is straightforward in the HC condition as the conversation is a continuation of the first half. However, for the LC condition, a completely new conversation is presented and sufficient dialogue must be attended to before the new themes of the video are established. Therefore, we suggest that in the HC condition, all of the elements necessary to construct a situation model are available to participants within the first few seconds of the videos, whereas in the LC condition, establishment of the thematic content can only be accomplished after there has been sufficient dialogue between the characters. While this interpretation is consistent with the ideas of Bar, Aminoff and colleagues (Bar and Aminoff 2003; Bar et al. 2008; Aminoff et al. 2013) and Ranganath and Ritchey (2012), it is important to note that other authors have proposed a more specialised role for the PHG/RSC in representing space (Epstein 2008; e.g. Mullally and Maguire 2011), and further work will be required to elucidate the exact contribution of these regions to processing of complex ongoing stimuli such as the video clips used in this study.

In contrast to the PHG and RSC, all other regions where activity was modulated by prior contextual information showed increases in the HC condition compared with the LC condition. In particular, the MTG, AG and IFG all showed a qualitatively similar response pattern – most clearly present in the MTG – where the activity to both HC and LC conditions increased after the video onset, but the activity in the HC condition peaked after around 10 seconds whereas the activity in the LC condition peaked after around 16 seconds (see Fig. 3). The HC condition elicited higher overall activity compared to the LC condition and this difference was most apparent around 8-10 seconds after the video started.

Within the MTG, there was not only a main effect of condition but a significant condition-by-time interaction, which was driven by the quicker time to peak in the HC condition (see Fig. 3 as well as Supplementary Fig. 4 for the locations showing the main and interaction effects). Given that the only difference between the HC and LC conditions is the prior knowledge about the theme of the narrative, what is the mechanism whereby these regions are most active in the HC condition?

We propose that the HC-specific effects in the MTG, AG and IFG reflect reactivations of a set of links between semantic concepts which are key to the understanding of the specific narrative contents and that can only have been established during viewing of the first half of the video. As an example, in the first half of one stimulus (taken from the 80s sitcom *Perfect Strangers*) the main character Larry explains to his roommate Balki (a recent immigrant from the fictional island of Mypos) his desire for the team he manages to win a baseball trophy. Unfortunately, his best player “Slugger” is injured. In the second half of the clip Balki offers to fill in for Slugger. Larry is forced to explain to him that baseball is difficult and without experience he won’t be able to fill Slugger’s shoes. For a participant in the LC condition, the characters and scene (i.e., the living room in which both scenes from this show used as stimuli were filmed) are familiar, but because they have not seen the first half of the stimulus the specific conceptual relations *between* baseball, coaching, trophies have not been established, and it is the reactivation of these links that is reflected in the activation increases observed in the HC condition. The fact that the difference between the HC and LC conditions are consistently larger during the early parts of the videos, as well as the significant time-by-condition interaction in the MTG,

leads us to hypothesize that the prior activation of these concepts and the links between them allows them to be activated sooner in the HC condition. A similar account was offered by Humphries et al. (2007), who proposed that the mid-MTG is involved in dynamic combinatorial semantic processing. For instance (in the example used by these authors), in the sentence “The man on vacation lost a bag and a wallet”, the specific meaning of the word ‘bag’ in the sentence is strongly constrained by the other elements, and in fact these constraints give the representation increased semantic detail. These authors speculated that the mid-MTG plays a key role in incorporating these constraints into the constantly developing semantic construct being created as the sentence unfolds, a hypothesis that dovetails nicely with the results presented here (see also Raposo et al. 2006)

Our suggestion that activation increases in the HC condition in the MTG, IFG and AG reflect activation of concepts necessary to comprehend the unfolding narratives does not rule out an explanation based on episodic retrieval of the first half of the videos. It is possible that the increased similarity of the HC videos to their corresponding first halves cues their retrieval. While this may be happening, there are a number of reasons why we have interpreted the effects in terms of conceptual activation. First, all three of these regions are strongly associated with semantic processing. To illustrate this, Supplementary Figure 7 shows the overlap between the areas exhibiting greater HC than LC activation, an influential meta-analysis of semantic processing (Binder et al., 2009) and a functional connectivity meta-analysis (Yeo et al. 2014) implemented in Neurosynth (neurosynth.org; Yarkoni et al. 2011) generated from the center of mass of our MTG cluster, which shows the same IFG and AG regions are part of a

larger functionally connected network of regions. Second, damage to these regions does not result in a significant loss of episodic memory but does result in semantic memory impairments (DeLeon et al. 2007; Schwartz et al. 2011; Robson et al. 2017). Third, there was no hint of a similar response pattern within regions more typically associated with episodic memory processing, such as the medial temporal lobes. Fourth, if the responses in these regions simply reflected the retrieval of the corresponding first half video, it is unclear why activity should extend beyond the first few seconds. Nevertheless, our task likely depends on the flexible interaction between both the episodic and semantic memory systems and further work will doubtless be required to tease apart the relative contributions of each. We will now consider the processing roles of these regions in greater detail.

Effectively the entire length of the left MTG is involved in semantic processing (Vigneau et al. 2006; Binder et al. 2009; Visser et al. 2012). The middle portion of the MTG/STS identified in our GLM analysis is most often implicated in lexico-semantic processing (Wible et al. 2006; Humphries et al. 2007), and its activation tracks the amount of semantic information in the stimulus (Humphries et al. 2007). Anatomically, it is important to distinguish the section of MTG discussed here from the putative semantic control region in the posterior MTG that has been studied extensively by Jefferies and colleagues (Jefferies and Lambon Ralph 2006; Whitney et al. 2011; Noonan et al. 2013), and identified in other work as well (e.g., Badre et al. 2005). The center of mass of our focus lies 1.5-2 cm anterior to the region of posterior MTG identified by Whitney et al. (2011) and Noonan et al. (2013). However, no significant voxels

were identified in this posterior region for the contrast of HC>LC, even at a relaxed threshold of $p < .05$.

Left IFG activation has consistently been observed in fMRI studies of semantic processing, though the degree to which these activations reflect an essential contribution to conceptual knowledge remains a subject of some debate. Binder et al. (2009), on the basis of their meta-analysis of neuroimaging studies of semantic processing, proposed that semantic processing in this structure is limited to its anterior and ventral aspects (specifically the pars orbitalis, BA 47). The focus we identify lays almost entirely in the ventral aspect of the pars triangularis (BA 45), extending into the lateral aspect of the pars orbitalis (BA 47), suggesting a role for this region in processing and reactivating the conceptual content of the HC videos (note as well that our focus lies almost entirely within the IFG region identified in Binder et al. (2009): see Supplementary Figure 8).

We also observed two activations in the lateral parietal lobe, one of which was located predominantly in the ventral AG, and the other, more dorsal, extending into the SMG. As can be seen in Figure 1, though these two clusters are quite close to one another, their response characteristics with regard to rest are quite different. While the more ventral cluster shows an increase in activation for videos compared to rest, the dorsal cluster shows the opposite pattern, with a general decrease triggered by the start of a stimulus. However, in both cases HC videos elicited stronger responses than LC videos.

The exact role of the AG has been the subject of recent debate in the literature, and this is further complicated by the functionally heterogeneity of this structure. Seghier et al. (2010) have proposed a tripartite division in which

the middle AG plays a constant role in conceptual processing while the dorsomedial and ventral AG mediate task-directed semantic processing. The focus we report here is closest to their middle AG region and extends into their ventral AG region, though as the coordinates reported for these two aspects of the AG are only 8 mm apart it is difficult to localize our finding to one area or the other given the 3 mm³ resolution of our data and the effects of smoothing. Finally, we observed a small cluster exhibiting an interaction between HC and LC in the right cuneus, which appeared to result from effects in the latter part of the video. Because of the unexpected nature of this finding we do not interpret it further here.

To conclude: though some might argue that television is an unsatisfactory substitute for real life, within the strictures imposed by fMRI it provides a very effective stand-in. Through manipulating the availability of knowledge about the narrative content of events depicted in short video clips, and investigating how brain activity changes during the course of the videos, we have delineated two sets of brain regions where brain activity is modulated in very different ways.

References

- Allen EA, Erhardt EB, Calhoun VD. 2012. Data visualization in the neurosciences: overcoming the curse of dimensionality. *Neuron*. 74:603–608.
- Aminoff EM, Kveraga K, Bar M. 2013. The role of the parahippocampal cortex in cognition. *Trends Cogn Sci*. 17:379–390.
- Badre D, Poldrack RA, Paré-Blagoev EJ, Insler RZ, Wagner AD. 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*. 47:907–918.
- Bar M, Aminoff E. 2003. Cortical analysis of visual context. *Neuron*. 38:347–358.
- Bar M, Aminoff E, Schacter DL. 2008. Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J Neurosci*. 28:8539–8544.
- Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where Is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex*. 19:2767–2796.
- Bird CM, Keidel JL, Ing LP, Horner AJ, Burgess N. 2015. Consolidation of complex events via reinstatement in posterior cingulate cortex. *J Neurosci*. 35:14426–14434.
- Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U. 2016. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci*. 20:115–125.
- Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 29:162–173.
- DeLeon J, Gottesman RF, Kleinman JT, Newhart M, Davis C, Heidler-Gary J, Lee A, Hillis AE. 2007. Neural regions essential for distinct cognitive processes

- underlying picture naming. *Brain*. 130:1408–1422.
- Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci*. 12:388–396.
- Epstein R, Graham KS, Downing PE. 2003. Viewpoint-specific scene representations in human parahippocampal cortex. 37:865–876.
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature*. 392:598–601.
- Esterman M, Tamber-Rosenau BJ, Chiu Y, Yantis S. 2010. Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage*. 50:572–576.
- Hasson U, Furman O, Clark D, Dudai Y, Davachi L. 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*. 57:452–462.
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science*. 303:1634–1640.
- Humphries C, Binder JR, Medler DA, Liebenthal E. 2007. Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage*. 36:924–932.
- Jefferies E, Lambon Ralph MA. 2006. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain*. 129:2132–2147.
- Maguire EA. 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand J Psychol*. 42:225–238.
- Mullally L, Maguire EA. 2011. A new role for the parahippocampal cortex in representing space. 31:7441–7449.

- Noonan KA, Jefferies E, Visser M, Ralph MAL. 2013. Going beyond inferior prefrontal involvement in semantic control : evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J Cogn Neurosci*. 25:1824–1850.
- Ranganath C, Ritchey M. 2012. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci*. 13:713–726.
- Raposo A, Moss HE, Stamatakis E, Tyler LK. 2006. Repetition suppression and semantic enhancement: an investigation of the neural correlates of priming. *Neuropsychologia*. 44:2284–2295.
- Robson H, Specht K, Beaumont H, Parkes LM, Sage K, Lambon Ralph MA, Zahn R. 2017. Arterial spin labelling shows functional depression of non-lesion tissue in chronic Wernicke’s aphasia. *Cortex*. 92:249–260.
- Schlochtermeyer LH, Pehrs C, Bakels J-H, Jacobs AM, Kappelhoff H, Kuchinke L. 2016. Context matters: Anterior and posterior cortical midline responses to sad movie scenes. *Brain Res*. 1661:24–36.
- Schwartz MF, Kimberg DY, Walker GM, Brecher A, Faseyitan OK, Dell GS, Mirman D, Coslett HB. 2011. Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc Natl Acad Sci*. 108:8520–8524.
- Seghier ML, Fagan E, Price CJ. 2010. Functional subdivisions in the left angular gyrus: where the semantic system meets and diverges from the default network. *J Neurosci*. 30:16809–16817.
- St-Laurent M, Abdi H, Buchsbaum BR. 2015. Distributed patterns of reactivation predict vividness of recollection. *J Cogn Neurosci*. 27.
- Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain.

Theime, Stuttgart, Ger.

- van Kesteren MTR, Fernandez G, Norris DG, Hermans EJ. 2010. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci.* 107:7550–7555.
- Vann SD, Aggleton JP, Maguire EA. 2009. What does the retrosplenial cortex do ? *Nat Rev Neurosci.* 10:792–802.
- Vigneau M, Beaucoisin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage.* 30:1414–1432.
- Visser M, Jefferies E, Embleton K V, Lambon Ralph MA. 2012. Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J Cogn Neurosci.* 24:1766–1778.
- Whitney C, Kirk M, Sullivan JO, Ralph MAL, Jefferies E. 2011. The neural organization of semantic control : TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb Cortex.* 21:1066–1075.
- Wible CG, Han SD, Spencer MH, Kubicki M, Niznikiewicz MH, Jolesz FA, McCarley RW, Nestor P. 2006. Connectivity among semantic associates: an fMRI study of semantic priming. *Brain Lang.* 97:294–305.
- Yarkoni T, Poldrack R, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods.*

8:665–870.

Yeo BTT, Krienen FM, Chee MWL, Buckner RL. 2014. Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *Neuroimage*. 88:212–227.

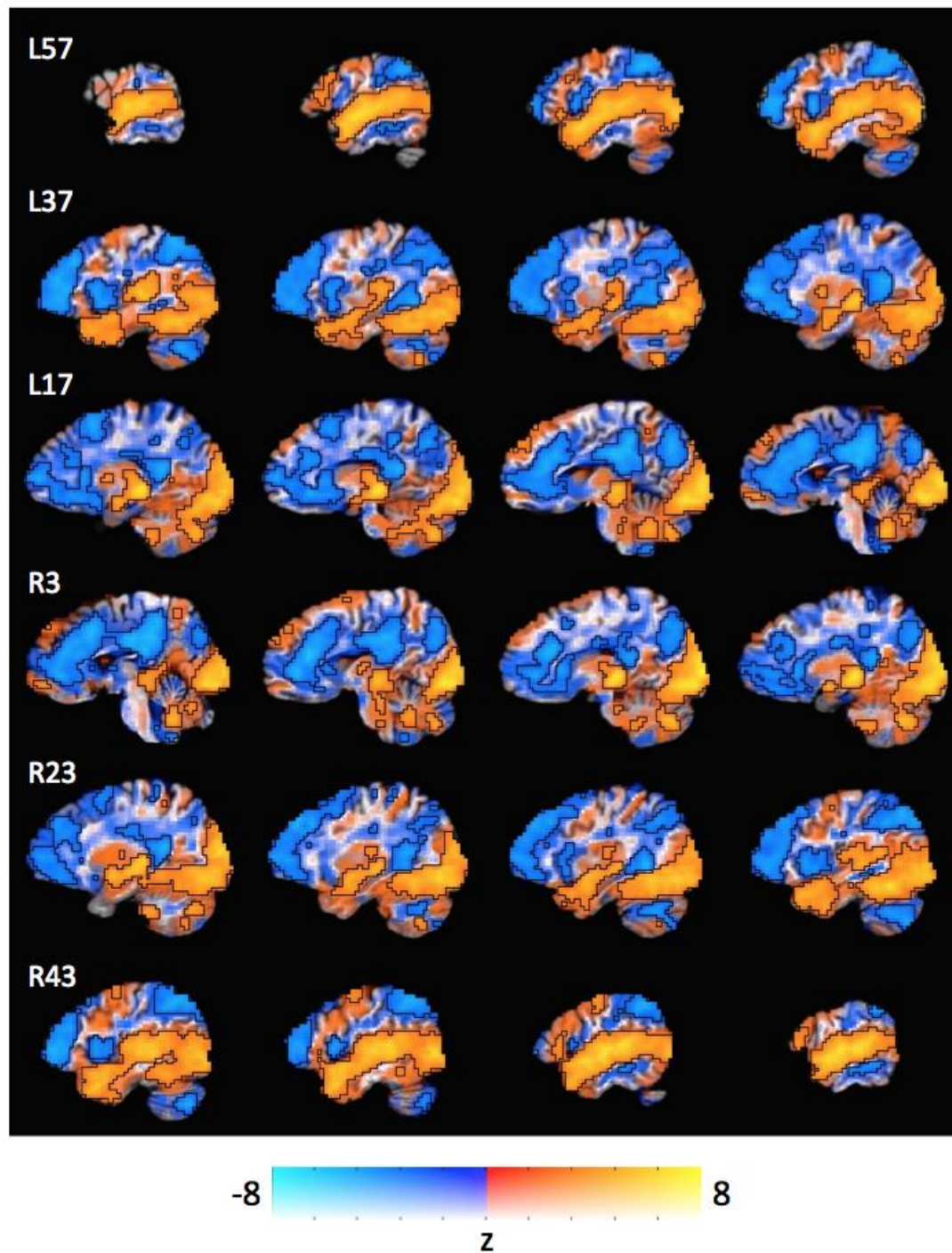
Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, Raichle ME. 2001. Human brain activity time-locked to perceptual event boundaries. *Nat Neurosci*. 4:651–655.

Zwaan R, Radvansky G. 1998. Situation models in language comprehension and memory. *Psychol Bull*. 123:162–185.

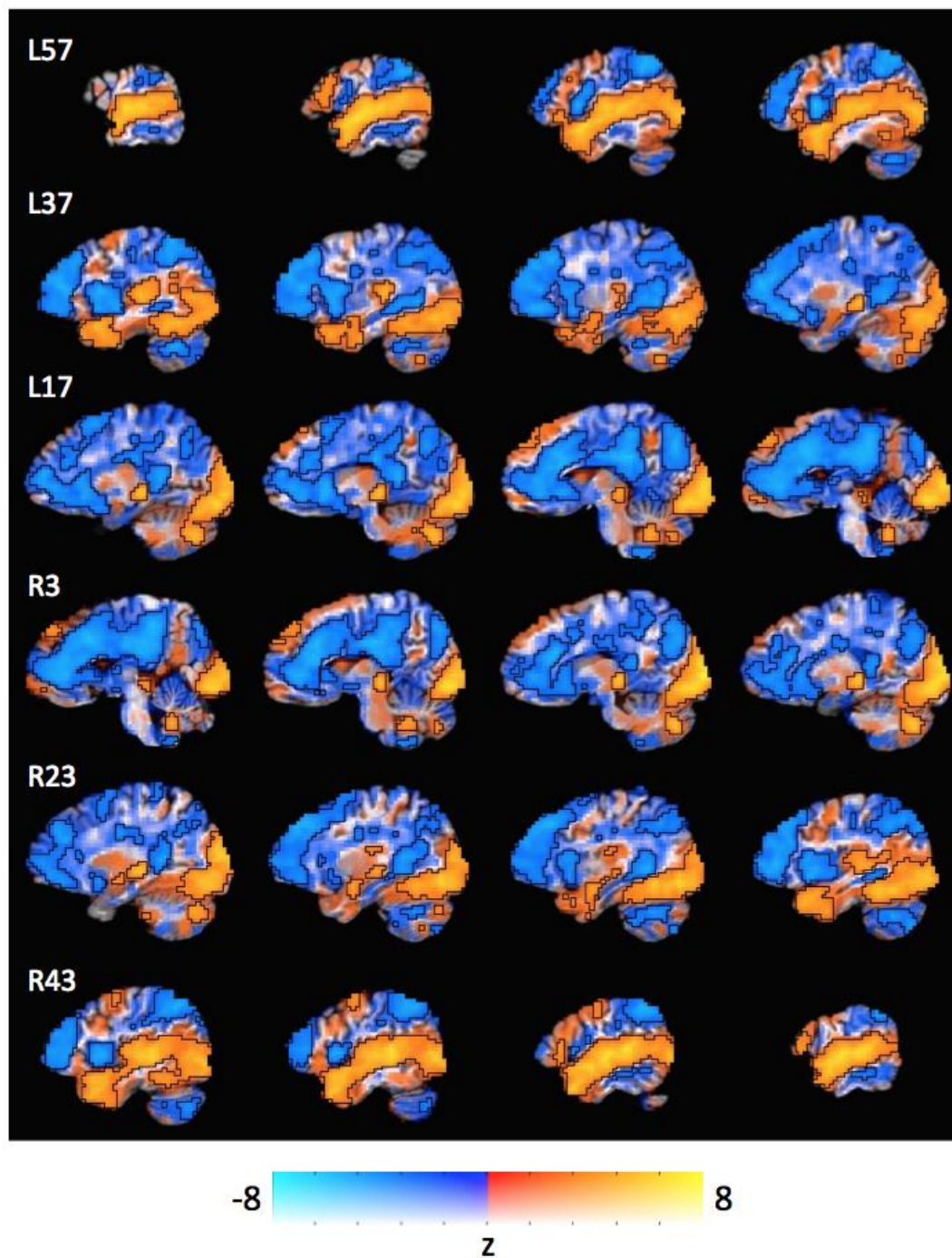
Region	# Voxels	CM x	CM y	CM z	Equivalent z at peak
<i>Main effect of condition</i>					
R retrosplenial	77	13	-49	11	-5.02
L middle temporal gyrus	57	-55	-29	1	4.75
L parahippocampal gyrus	42	-25	-45	-6	-5.39
L angular gyrus	38	-52	-54	25	4.59
L inferior frontal gyrus (BA 45/47)	25	-50	15	2	4.63
L supramarginal gyrus	25	-49	-54	39	4.44
<i>Interaction</i>					
L middle temporal gyrus	30	-52	-31	3	4.69
R cuneus	19	8	-71	8	4.62

Table 1. Significant effects identified in Condition (HC vs. LC) by Time ANOVA. F-statistics have been converted to z-statistics via their respective cumulative distribution functions.

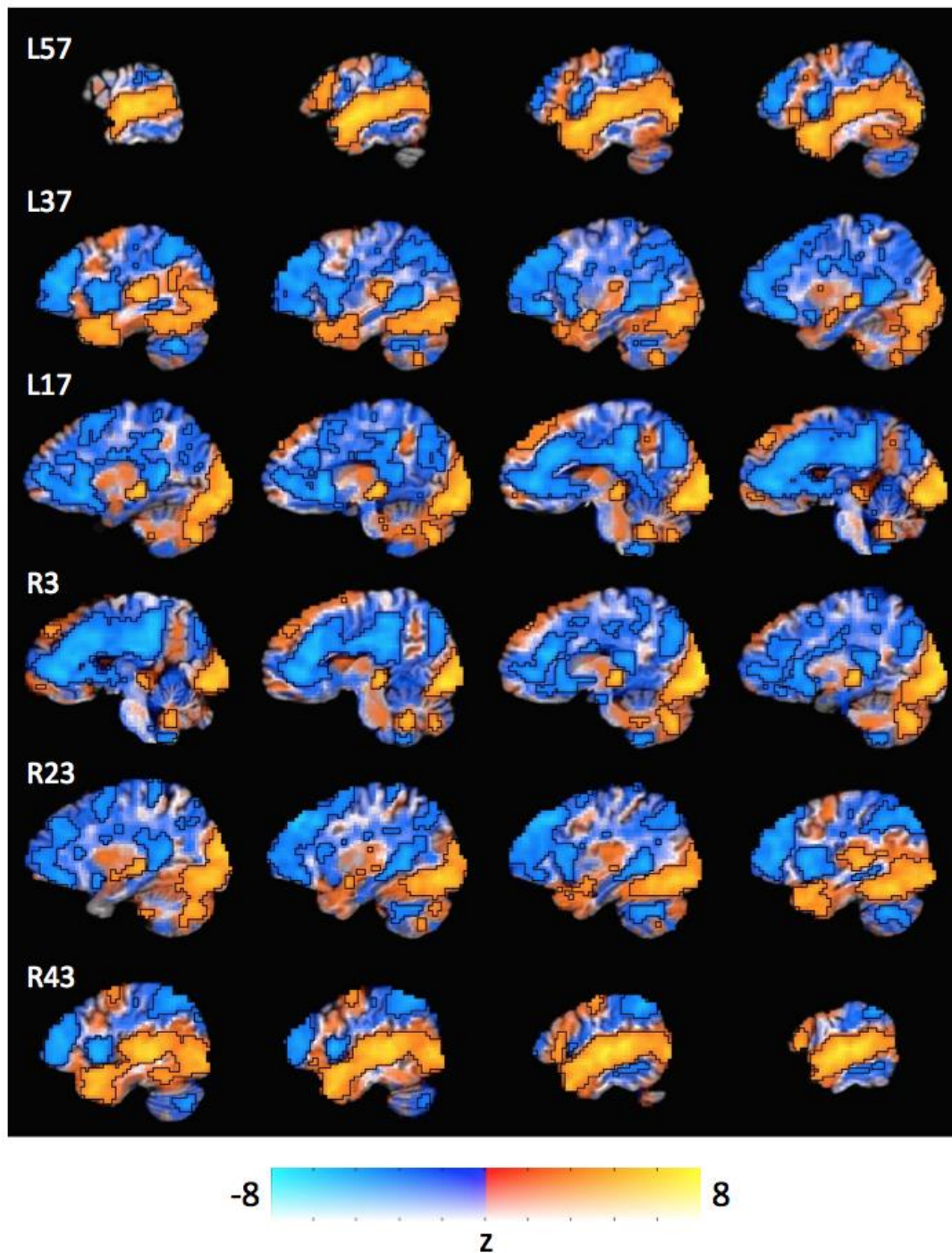
Supplementary Figures



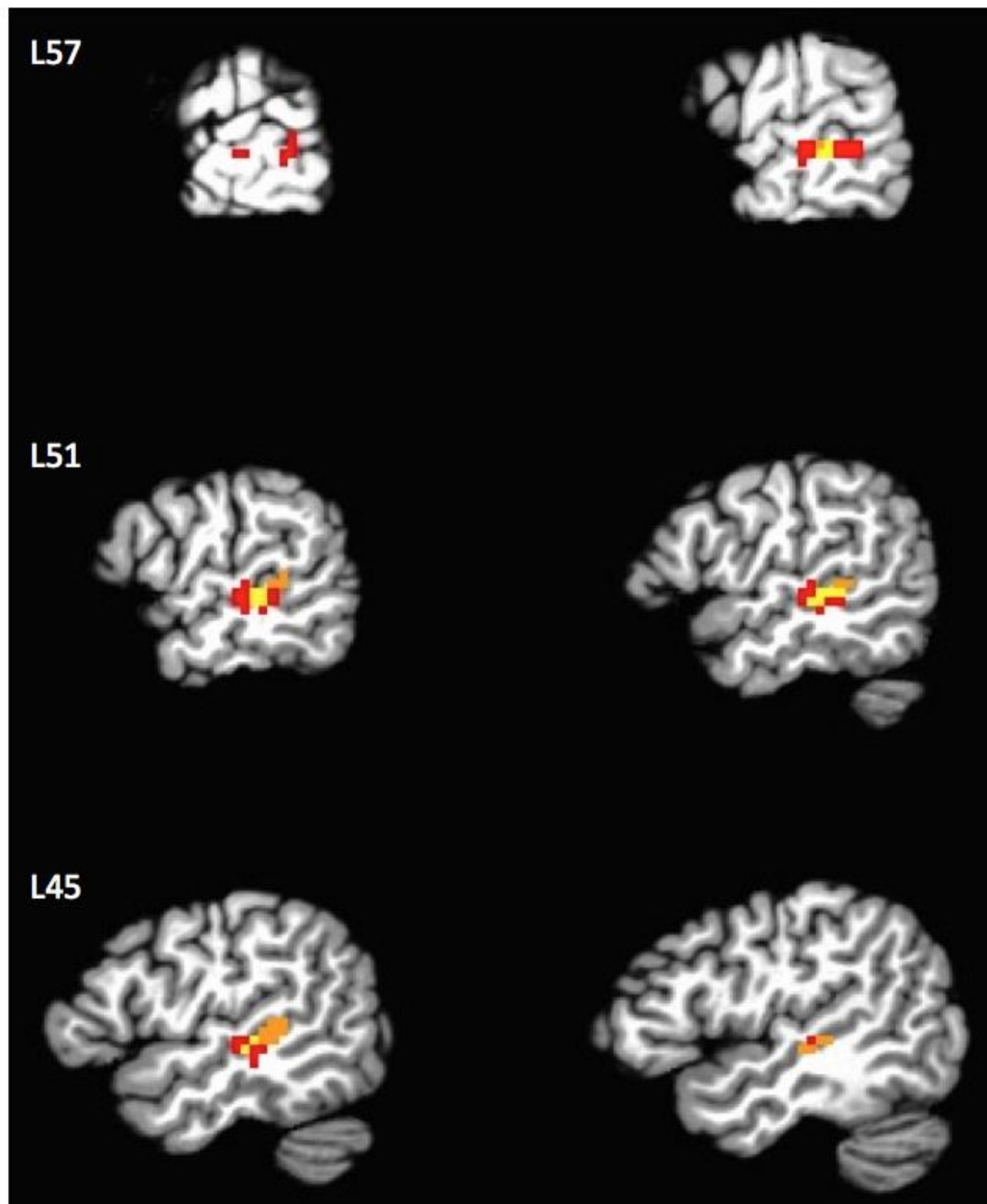
Supplementary Figure 1. Contrast of NC vs. rest using block design (see Methods) showing areas that are active (warm colors) or deactive (cool colors) during the first half video clips relative to the resting baseline. Areas outlined in black are significant at $p < .001$ uncorrected. Betas are mapped to hue and t-statistics to transparency, following Allen et al. (2012). Slices are at 5 mm intervals along the x-axis from left to right.



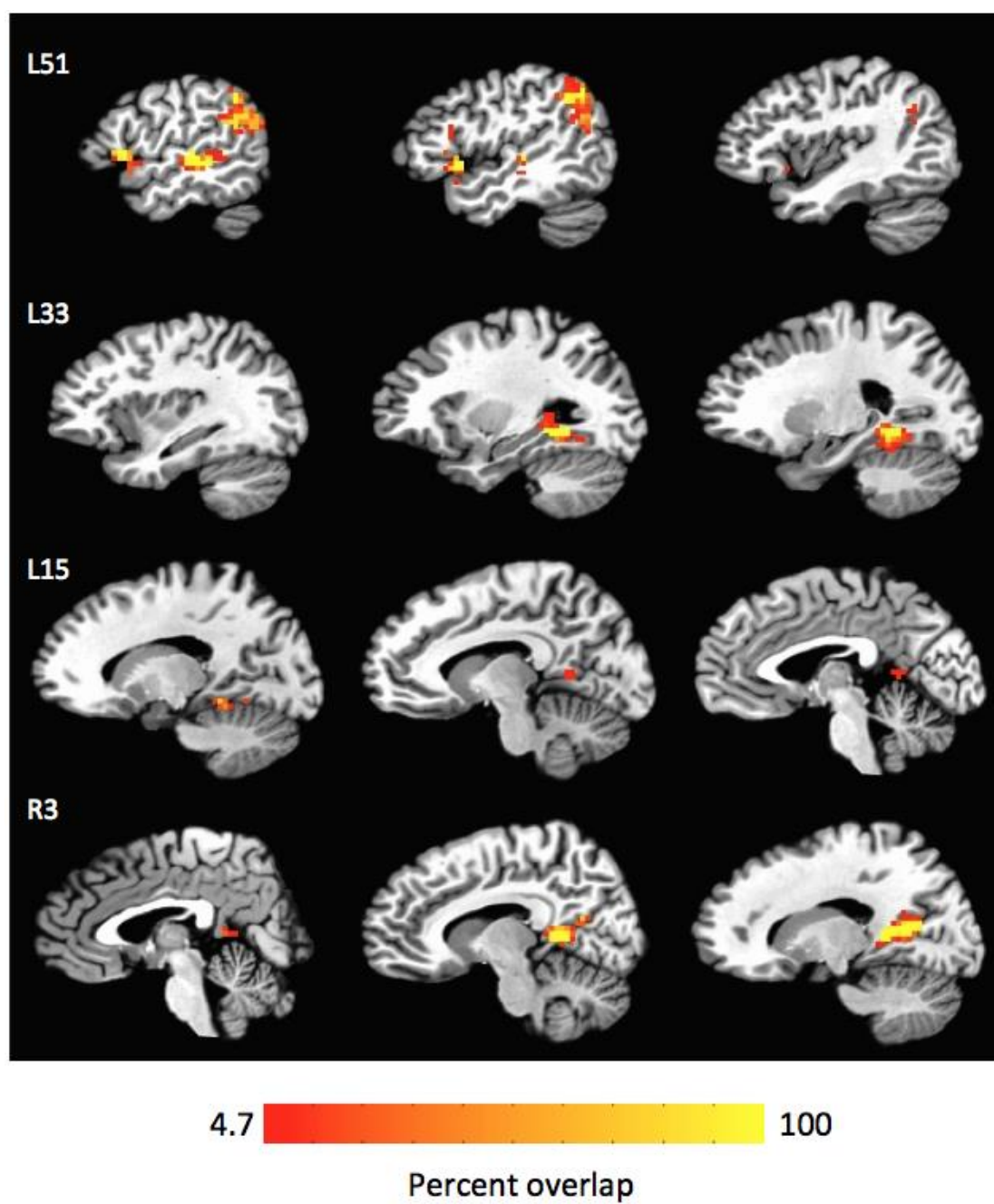
Supplementary Figure 2. Contrast of LC vs. rest using block design (see Methods) showing areas that are active (warm colors) or deactive (cool colors) during the second half non-continuation video clips relative to the resting baseline. Areas outlined in black are significant at $p < .001$ uncorrected. Betas are mapped to hue and t-statistics to transparency, following Allen et al. (2012). Slices are at 5 mm intervals along the x-axis from left to right.



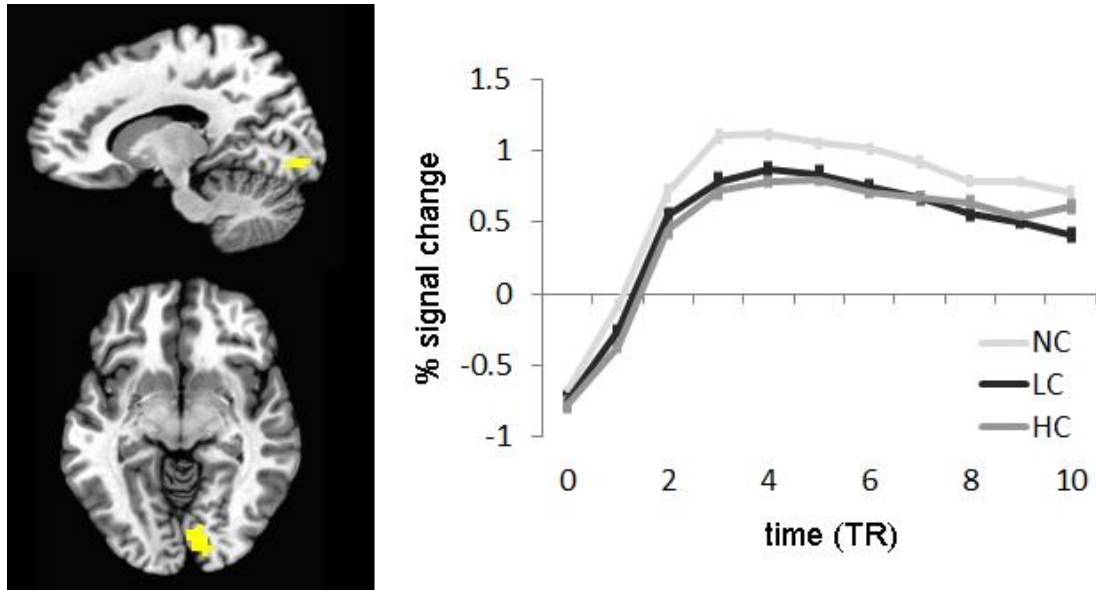
Supplementary Figure 3. Contrast of HC vs. rest using block design (see Methods) showing areas that are active (warm colors) or deactive (cool colors) during the second half continuation video clips relative to the resting baseline. Areas outlined in black are significant at $p < .001$ uncorrected. Betas are mapped to hue and t-statistics to transparency, following Allen et al. (2012). Slices are at 5 mm intervals along the x-axis from left to right.



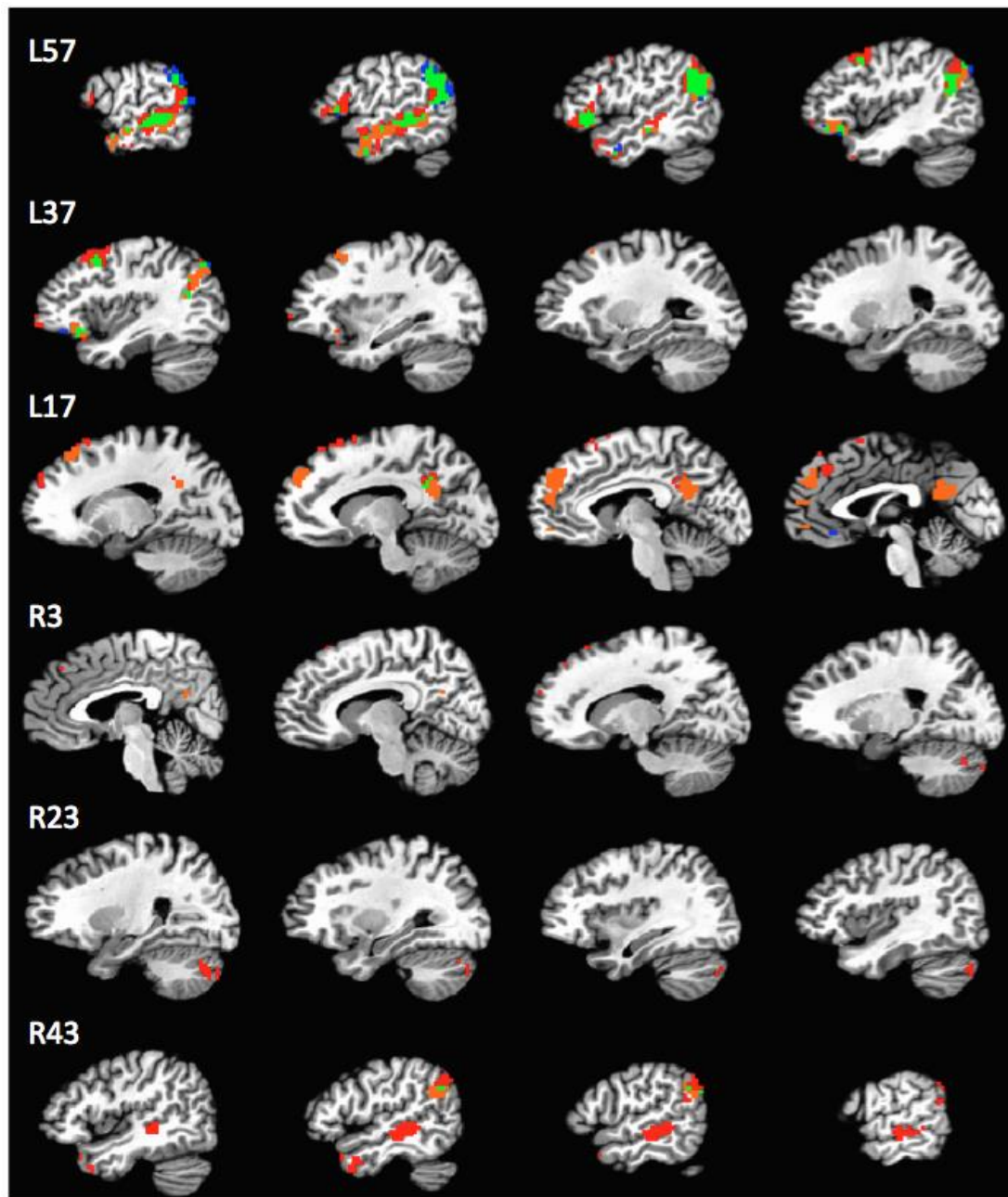
Supplementary Figure 4. Overlap of MTG clusters obtained in the main effect and interaction of condition x time ANOVA. Voxels in red showed a whole-brain significant main effect, voxels in orange showed a whole-brain significant interaction, and voxels in yellow showed both effects.



Supplementary Figure 5. Overlap of LOSO-defined clusters.



Supplementary Figure 6. Time course of activation in a representative cluster in early visual cortex. Cluster center of mass located at +10, -81, -4.



Supplementary Figure 7. Conjunction of HC > LC contrast (thresholded at $p < .05$ uncorrected), functional connectivity derived from neurosynth.org (Yarkoni et al. 2011) using the center of mass of the MTG cluster $[-59, -31, -1]$ as the seed and thresholded at default of $r = 0.2$, and semantic network from meta-analysis in Binder et al. (2009). Voxels in green represent overlap of all three maps. Voxels in blue represent overlap of current results and Binder et al. map. Voxels in red represent overlap of our results and functional connectivity map from Neurosynth. Voxels in orange represent overlap of Binder et al. map and Neurosynth map. This shows that the regions where we found HC>LC effects are part of a larger functional network and are part of the semantic network.

Supplementary information: Appendix A. Details of the clips used in the experiment

Show

The Bob Newhart Show (living room couch)

Season 1, Episode 5: Bob's ex-wife has called, and he discusses this with Emily

Season 1, Episode 11: Bob tells Emily that he needs time to himself

Butterflies (kitchen table)

Season 3, Episode 6: Ben is going to a dentist's convention

Season 3, Episode 7: Ben complains about electricity bill

Curb Your Enthusiasm (in a restaurant)

Season 3, Episode 5: The Davids dine with a couple and discuss Alanis Morissette

Season 3, Episode 7: The Davids dine with same couple and discuss getting a dog

Dharma and Greg (kitchen)

Season 2, Episode 14: Greg has a hangover

Season 2, Episode 21: Greg defends his mother to Dharma

Ever Decreasing Circles (kitchen)

Season 2, Episode 2: Martin is getting ready for cricket match

Season 2, Episode 3: Martin does not want to have dinner with new people

Family Ties (living room)

Series 1, Episode 1: Steven calls country club to get Alex to come home

Series 1, Episode 5: Steven has issues with his father

The Golden Girls (kitchen table)

Series 2, Episode 13: Rose tells story about a sheep she once knew

Series 2, Episode 15: Blanche describes a date she has been on

Growing Pains (kitchen)

Series 1, Episode 5: Jason and Maggie discuss baby names

Series 1, Episode 15: Jason and Maggie discuss raising their daughter

Just Shoot Me (Jack's office)

Series 2, Episode 10: Maya and Jack discuss the importance of confidence

Series 2, Episode 24: Maya and Jack discuss her mother/his ex-wife

Mad About You (living room)

Season 2, Episode 15: Paul and Jamie discuss why she is still mad at him

Season 2, Episode 22: Paul tells Jamie he has seen a mutual friend

Moonlighting (Maddie's office)

Series 2, Episode 9: Maddie tells David she is going to sell the agency

Series 2, Episode 16: Maddie and David celebrate a work anniversary

Magnum, P.I. (Magnum's study)

Series 4, Episode 6: Magnum and Higgins discuss the kidnapped duchess

Series 4, Episode 9: Higgins has a cold; Magnum tries to cheer him up

Northern Exposure (Maggie's living room)

Season 2, Episode 2: Maggie and Joel discuss her feelings for him

Season 2, Episode 2: Maggie and Joel discuss the shrines she has made to dead exes

NewsRadio (Dave's office)

Series 3, Episode 11: Dave and Lisa discuss his being secretly Canadian

Series 3, Episode 23: Dave and Lisa discuss a terrible interview he's given

The Odd Couple (living room)

Season 2, Episode 12: Felix and Oscar argue about who gets to use the apartment

Season 2, Episode 17: Felix and Oscar discuss cleanliness of windows

Perfect Strangers (kitchen/living room)

Season 2, Episode 3: Balki and Larry discuss baseball tournament

Season 2, Episode 16: Balki and Larry discuss tuxedos

Seinfeld (kitchen)

Season 5, Episode 13: Jerry and Elaine mock George's new winter coat

Season 5, Episode 14: Jerry and Elaine laugh at George's Jack Nicholson voice

Three's Company (living room couch)

Season 2, Episode 2: Janet and Chrissy try to help Jack find a job

Season 2, Episode 3: Janet discusses a promotion with Jack and Chrissy

Whatever Happened to the Likely Lads? (the pub—same table)

Season 2, Episode 2: Bob tells Terry he needs to get his life together

Season 2, Episode 3: Terry tells Bob about saying goodbye to his girlfriend

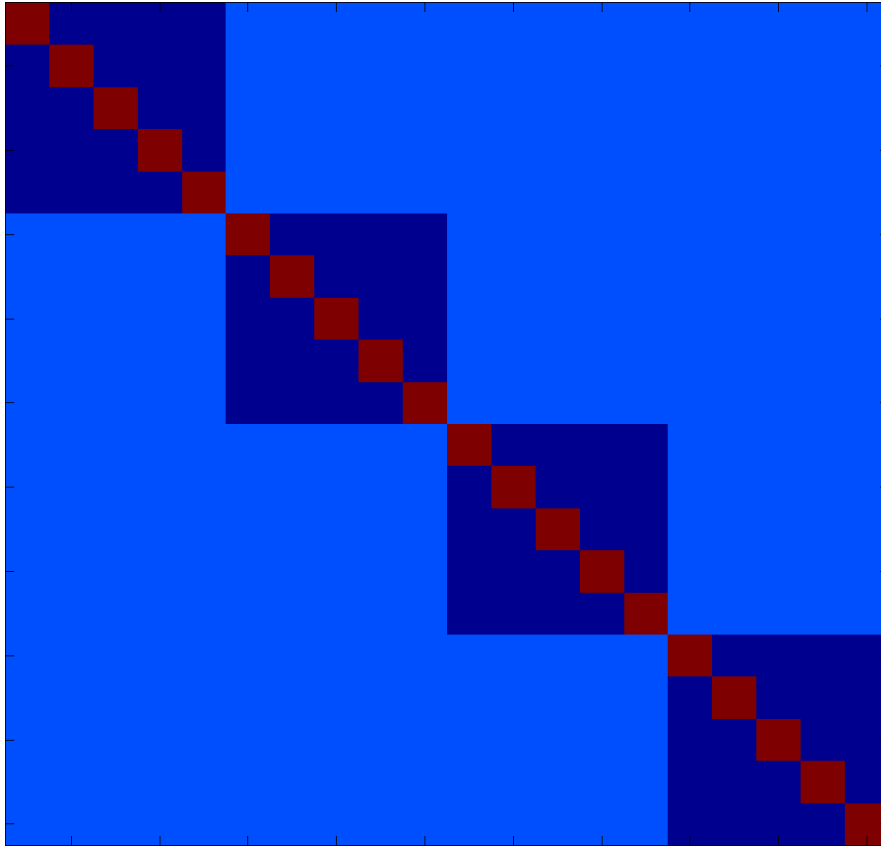
Who's The Boss? (living room)

Season 1, Episode 1: Angela is mad at Tony for hanging out with her boyfriend

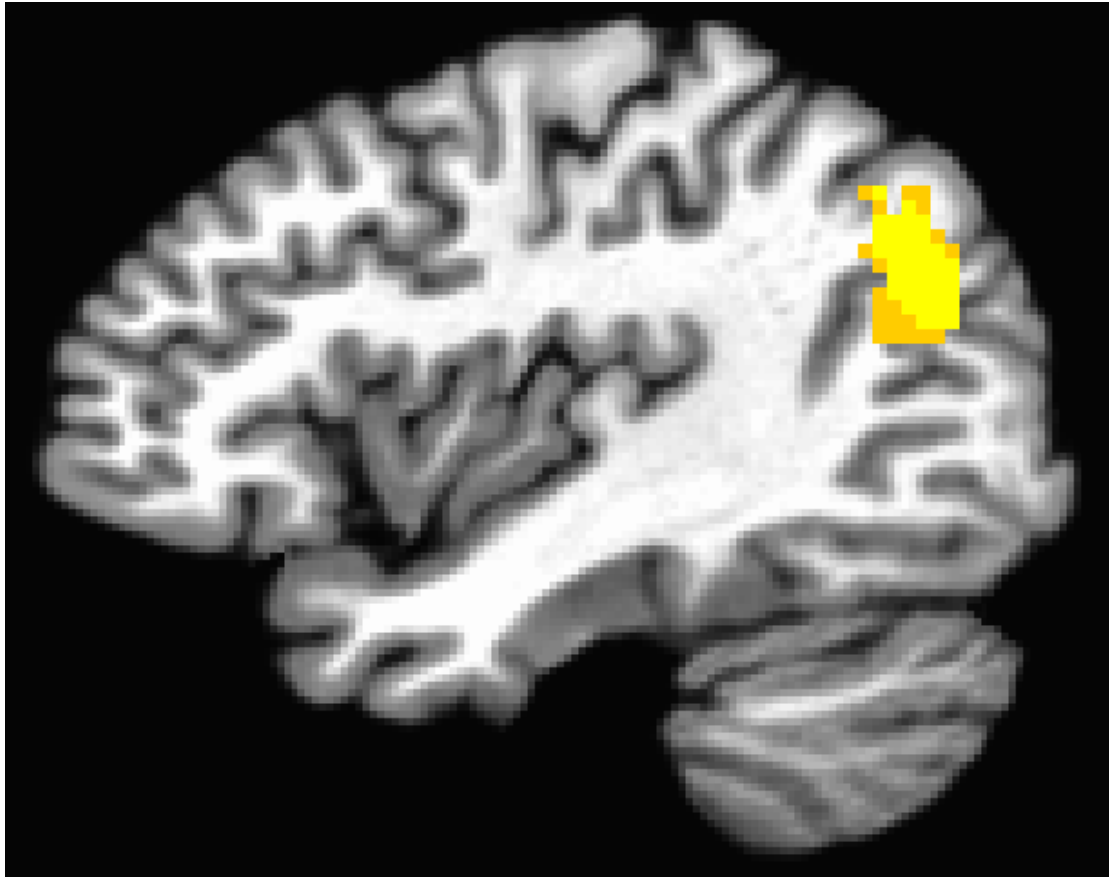
Season 1, Episode 9: Angela and Tony discuss a birthday present for Sam

Supplementary Information: Appendix B. Details of the representational similarity analysis.

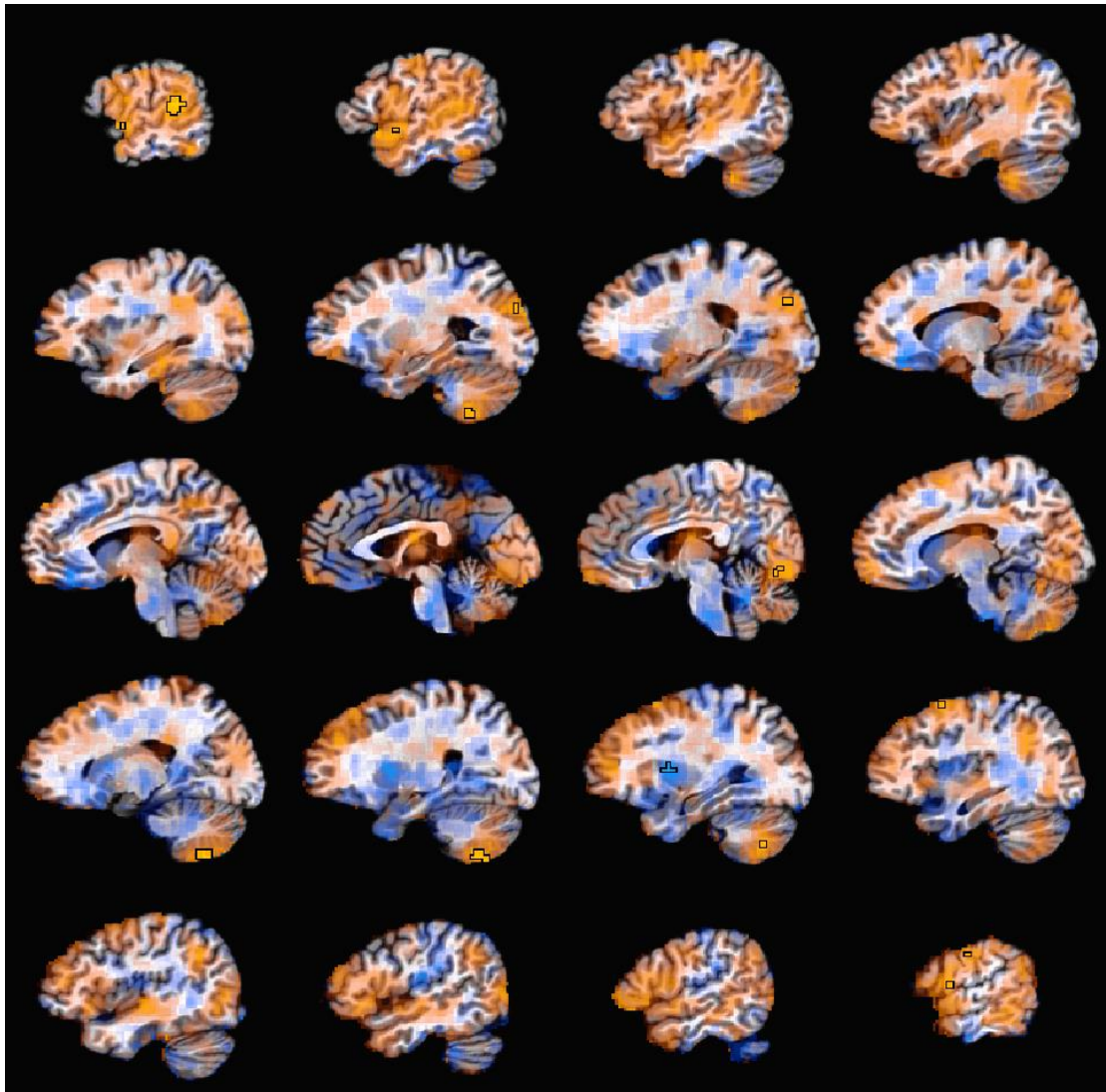
MVPA Analysis. To identify regions exhibiting representational similarity between clips taken from the same show irrespective of condition (LC or HC), constructed a contrast matrix with large positive values on the diagonal (corresponding to clips taken from the same show) and small negative values off the diagonal (corresponding to clips taken from different shows), such that the sum of these constraints was zero. An additional constraint on the formation of this matrix comes from the fact that videos were presented in groups of five first halves and five second halves. Because temporal proximity of stimuli can act as a confound in MVPA analyses (specifically, activation patterns of temporally close stimuli are more similar to one another, all other things being equal), we set all matrix cells that compared videos from different sets of first halves and second halves to zero. By design, this includes only non-matching pairs. The matrix for this contrast is depicted below—red squares represent matching first and second halves (positively weighted), dark blue squares represent temporally proximate non-matching first and second halves (negatively weighted), and light blue squares represent non-matching, non-temporally proximate video pairs (these values were set to 0 and therefore do not play a part in the analysis).



We calculated the matrix of all Fisher-Z transformed correlations for each video pair and multiplied them by the above contrast matrix, then assigned the resulting value to the center of the searchlight. We then used a one-sample t-test against 0 to assess whole-brain significant group-level effects for this contrast. This analysis yielded a single focus in the left angular gyrus (center of mass: -38, -65, 26), located posterior and dorsal to the angular gyrus focus reported from the GLM analysis in the main text:



We then tested for regions whose similarity between clips taken from the same show was modulated by condition (HC or LC). To this end, we weighted on-diagonal correlations corresponding to the correlation between an NC and HC video with +1 and NC-LC pairs with a -1, and assigned the inner product of this vector and the vector of on-diagonal correlations above to the center of each searchlight. We then tested the resulting maps with a one-sample t-test against 0. No significant clusters were identified in this analysis. However, we present below an unthresholded map wherein voxels with an uncorrected $p < .01$ are outlined in black according to the guidelines introduced by Allen et al. (2012).



References

- Allen EA, Erhardt EB, Calhoun VD. 2012. Data Visualization in the Neurosciences: Overcoming the Curse of Dimensionality. *Neuron*. 74:603–608.
- Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cereb Cortex*. 19:2767–2796.
- Yarkoni T, Poldrack R, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*. 8:665–870.